

Appendix

Appendix A1 Study characteristics: Dynarski, Gleason, Rangarajan, & Wood, 1998 (randomized controlled trial with differential attrition)

Characteristic	Description
Study citation	Dynarski, M., Gleason, P., Rangarajan, A., & Wood, R. (1998). <i>Impacts of dropout prevention programs: Final report. A research report from the School Dropout Demonstration Assistance Program evaluation</i> . Princeton, NJ: Mathematica Policy Research, Inc.
Participants	<p>The <i>Twelve Together</i> study used a randomized controlled trial research design. The study sample of 500 students was comprised of two cohorts. Cohort 1 included 130 students in the intervention group and 116 students in the control group who were eighth graders during the 1992–93 school year. Cohort 2 included 133 students in the intervention group and 121 students in the control group who were eighth graders during the 1993–94 school year. Random assignment occurred at the beginning of the eighth grade.</p> <p>Participation in the program was voluntary. Participants were recruited through presentations at the schools and through referrals from school staff. To ensure a mix of academic abilities in the program, the schools divided applicants into high, medium, and low risk categories. Within these categories, schools paired similar students and ranked the pairs in terms of priority for entry into the program. Regardless of the number of applicants within each risk category, when filling a particular <i>Twelve Together</i> group, equal numbers of student pairs were chosen from each category, starting with the highest priority pair within each category. One student from each pair was then randomly selected to enter the group; the other was assigned to the control group. Therefore, each group included equal numbers of students from each of the three risk categories.</p> <p>Participants were, on average, just under 14 years old at the time of random assignment. About half of the sample was Hispanic; about a quarter was Asian or other ethnicities; about one in seven was white; and about one in ten was black. Just over a quarter of participants lived in one-parent or no-parent households. About one student in seven lived in a household receiving public assistance, and about one in five did not speak English at home. Slightly more than a third of participants had discipline problems during the previous school year, and about a quarter of students were not sure if they would finish high school.</p> <p>To ensure adequate follow-up for assessing effects on staying in school and progressing in school, the WWC used the third-year follow-up outcomes for determining the effectiveness rating for <i>Twelve Together</i>. These third-year outcomes are available for cohort 1 only. These results are drawn from a follow-up survey administered to students in cohort 1 in 1995, approximately three years after random assignment; 119 intervention group students and 100 control group students responded—for response rates of 92% and 86% respectively. Because these response rates represent differential attrition of more than five percentage points (the differential attrition threshold used for WWC dropout prevention reviews), the WWC rated this study as meeting evidence standards with reservations.</p> <p>Researchers compared the baseline characteristics of third-year follow-up respondents in the two research groups on 13 demographic, socioeconomic, and school performance measures. A statistical test of the overall difference between the research groups on the full set of 13 baseline characteristics found that the groups were not significantly different.</p>
Setting	The <i>Twelve Together</i> study was conducted in nine middle schools in the Sweetwater Union High School District located in Chula Vista, California (near San Diego).
Intervention	The <i>Twelve Together</i> program in Chula Vista, California, was a peer support and mentoring program for eighth graders. The one-year voluntary program offered weekly after-school discussion groups. Each group of about 12 students included equal numbers of students from each of three academic risk groups (high, medium, and low) as categorized by the school. Groups were led by two trained adult volunteer facilitators who moderated discussions. Topics, chosen based on student interest, focused on personal, family, and social issues. In addition to attending discussion sessions, participants agreed to study regularly, not to skip classes, and to work to improve their grades. Facilitators, usually college students, also provided homework assistance. To promote group cohesion and develop teamwork skills, the program began with a weekend camping outing. It also provided other activities such as visits to college campuses and social events. ¹

(continued)

Appendix A1

Study characteristics: Dynarski, Gleason, Rangarajan, & Wood, 1998
(randomized controlled trial with differential attrition) (continued)

Characteristic	Description
Comparison	Control group students did not participate in <i>Twelve Together</i> and attended the same middle schools as students in the intervention group. Analysis presented in the impact report for the study indicates that control group students were more likely than intervention group students to have participated in remedial classes in reading, math, and other subjects (41% compared with 32%).
Primary outcomes and measurement	Two relevant outcomes from the <i>Twelve Together</i> study are included in this summary: the dropout rate and highest grade completed. (See Appendices A2.1 and A2.2 for a more detailed description of these outcome measures.) The study also examined the program's effects on absenteeism, English and math grades, self-esteem, and perceived likelihood of completing high school. These outcomes do not fall within the three domains examined by the WWC's review of dropout prevention interventions (staying in school, progressing in school, and completing school) and are not included in this summary.
Teacher training	Most <i>Twelve Together</i> facilitators were college students or recent college graduates from San Diego State University. During the time the program was operating, the university offered a credit bearing course to train students to be <i>Twelve Together</i> facilitators.

1. This information is based on the implementation report from the study (Hershey, Adelman, & Murray, 1995), as well as additional details gathered from the study authors.

Appendix A2.1 Outcome measures for the staying in school domain

Outcome measure	Description
Dropped out of school	The percentage of students who had dropped out of high school at the end of the third follow-up year. These data were self-reported and were collected through follow-up surveys.

Appendix A2.2 Outcome measures for the progressing in school domain

Outcome measure	Description
Highest grade completed	The grade level completed by the end of the third follow-up year. These data were self-reported and were collected through follow-up surveys.

Appendix A3.1 Summary of study findings included in the rating for the staying in school domain¹

Outcome measure	Study sample	Sample size (schools/ students)	Author's findings from the study					
			Mean outcome		WWC calculations			
			Twelve Together group	Comparison group	Mean difference ² (Twelve Together – comparison)	Effect size ³	Statistical significance ⁴ (at $\alpha = 0.05$)	Improvement index ⁵
Dynarski et al., 1998 (randomized controlled trial with differential attrition) ⁶								
Dropped out of school (%)	Cohort 1	9/219	8	13	5	0.33	ns	+13
Domain average ⁷ for staying in school						0.33	ns	+13

ns = not statistically significant

1. This appendix reports findings considered for the effectiveness rating and the average improvement index. These results were measured at the end of the third year and are available for cohort 1 only. Second-year follow-up results in this domain, available for both cohorts 1 and 2, are not included in these ratings but are reported in Appendix A4.1.
2. Positive differences and effect sizes favor the intervention group; negative differences and effect sizes favor the comparison group. For the dropped out of school outcome, signs were reversed on the mean difference, effect size, and improvement index to demonstrate that the intervention group was favored when negative differences were reported. Means from Dynarski et al. (1998) are estimated using regression models that control for baseline characteristics.
3. Effect sizes for dichotomous variables were computed using the Cox Index. For an explanation of the effect size calculation, see [Technical Details of WWC-Conducted Computations](#).
4. Statistical significance is the probability that the difference between groups is a result of chance rather than a real difference between the groups.
5. The improvement index represents the difference between the percentile rank of the average student in the intervention condition and that of the average student in the comparison condition. The improvement index can take on values between –50 and +50, with positive numbers denoting results favorable to the intervention group.
6. The level of statistical significance was reported by the study authors or, where necessary, calculated by the WWC to correct for clustering within classrooms or schools and for multiple comparisons. For an explanation about the clustering correction, see the [WWC Tutorial on Mismatch](#). See [Technical Details of WWC-Conducted Computations](#) for the formulas the WWC used to calculate statistical significance. In the case of Dynarski et al. (1998), no corrections for clustering or multiple comparisons were needed.
7. This row provides the study average, which in this case is also the domain average. The WWC-computed domain average effect size is a simple average rounded to two decimal places. The domain improvement index is calculated from the average effect size.

Appendix A3.2 Summary of study findings included in the rating for the progressing in school domain¹

Outcome measure	Study sample	Sample size (schools/ students)	Author's findings from the study		WWC calculations			
			Mean outcome (standard deviation ²)		Mean difference ³ (<i>Twelve Together</i> – comparison)	Effect size ⁴	Statistical significance ⁵ (at $\alpha = 0.05$)	Improvement index ⁶
			<i>Twelve Together</i> group	Comparison group				
Dynarski et al., 1998 (randomized controlled trial with differential attrition) ⁷								
Highest grade completed	Cohort 1	9/219	9.1 (0.6)	9.2 (0.7)	−0.1	−0.15	ns	−6
Domain average ⁸ for progressing in school						−0.15	ns	−6

ns = not statistically significant

1. This appendix reports findings considered for the effectiveness rating and the average improvement index. These results were measured at the end of the third year and are available for cohort 1 only. Second-year follow-up results in this domain, available for both cohorts 1 and 2, are not included in these ratings but are reported in Appendix A4.2.
2. The standard deviation across all students in each group shows how dispersed the participants' outcomes are: a smaller standard deviation on a given measure would indicate that participants had more similar outcomes. Standard deviations for highest grade completed are not included in Dynarski et al. (1998) and were reported to the WWC by the study authors.
3. Positive differences and effect sizes favor the intervention group; negative differences and effect sizes favor the comparison group. Means from Dynarski et al. (1998) are estimated using regression models that control for baseline characteristics.
4. For an explanation of the effect size calculation, see [Technical Details of WWC-Conducted Computations](#).
5. Statistical significance is the probability that the difference between groups is a result of chance rather than a real difference between the groups. The WWC obtained additional information on the level of statistical significance of this difference from the study authors.
6. The improvement index represents the difference between the percentile rank of the average student in the intervention condition and that of the average student in the comparison condition. The improvement index can take on values between –50 and +50, with positive numbers denoting results favorable to the intervention group.
7. The level of statistical significance was reported by the study authors or, where necessary, calculated by the WWC to correct for clustering within classrooms or schools and for multiple comparisons. For an explanation about the clustering correction, see the [WWC Tutorial on Mismatch](#). See [Technical Details of WWC-Conducted Computations](#) for the formulas the WWC used to calculate statistical significance. In the case of Dynarski et al. (1998), no corrections for clustering or multiple comparisons were needed.
8. This row provides the study average, which in this case is also the domain average. The WWC-computed domain average effect size is a simple average rounded to two decimal places. The domain improvement index is calculated from the average effect size.

Appendix A4.1 Summary of shorter-term findings for the staying in school domain¹

Outcome measure	Study sample	Sample size (schools/ students)	Author's findings from the study					
			Mean outcome		WWC calculations			
			Twelve Together group	Comparison group	Mean difference ² (Twelve Together – comparison)	Effect size ³	Statistical significance ⁴ (at $\alpha = 0.05$)	Improvement index ⁵
Dynarski et al., 1998 (randomized controlled trial with differential attrition) ⁶								
Dropped out of school at end of second year (%)	Cohorts 1 and 2	9/466	7	8	1	0.09	ns	+3

ns = not statistically significant

1. This appendix presents second-year follow-up findings for measures in the staying in school domain. These results are available for both cohorts 1 and 2. To ensure an adequate follow-up period for a middle school dropout prevention program, the WWC used the third-year results to assess the effectiveness of *Twelve Together*. These results, available for cohort 1 only, are presented in Appendix A3.1.
2. Positive differences and effect sizes favor the intervention group; negative differences and effect sizes favor the comparison group. For the dropped out of school outcome, signs were reversed on the mean difference, effect size, and improvement index to demonstrate that the intervention group was favored when negative differences were reported. Means from Dynarski et al. (1998) are estimated using regression models that control for baseline characteristics.
3. Effect sizes for dichotomous variables were computed using the Cox Index. For an explanation of the effect size calculation, see [Technical Details of WWC-Conducted Computations](#).
4. Statistical significance is the probability that the difference between groups is a result of chance rather than a real difference between the groups.
5. The improvement index represents the difference between the percentile rank of the average student in the intervention condition and that of the average student in the comparison condition. The improvement index can take on values between –50 and +50, with positive numbers denoting results favorable to the intervention group.
6. The level of statistical significance was reported by the study authors or, where necessary, calculated by the WWC to correct for clustering within classrooms or schools (corrections for multiple comparisons were not done for findings not included in the overall intervention rating). For an explanation about the clustering correction, see the [WWC Tutorial on Mismatch](#). See [Technical Details of WWC-Conducted Computations](#) for the formulas the WWC used to calculate statistical significance. In the case of Dynarski et al. (1998), no corrections for clustering or multiple comparisons were needed.

Appendix A4.2 Summary of shorter-term findings for the progressing in school domain¹

Outcome measure	Study sample ⁴	Sample size (schools/ students)	Author's findings from the study		WWC calculations			
			Mean outcome (standard deviation ²)		Mean difference ³ (<i>Twelve Together</i> – comparison)	Effect size ⁴	Statistical significance ⁵ (at $\alpha = 0.05$)	Improvement index ⁶
			<i>Twelve Together</i> group	Comparison group				
Dynarski et al., 1998 (randomized controlled trial with differential attrition) ⁷								
Highest grade completed at end of second year	Cohorts 1 and 2	9/466	8.6 (0.5)	8.7 (0.5)	−0.1	−0.12	ns	−5

ns = not statistically significant

1. This appendix presents second-year follow-up findings for measures in the progressing in school domain. These results are available for both cohorts 1 and 2. To ensure an adequate follow-up period for a middle school dropout prevention program, the WWC used the third-year results for progressing in school to assess the effectiveness of *Twelve Together*. These results, available for cohort 1 only, are presented in Appendix A3.2.
2. The standard deviation across all students in each group shows how dispersed the participants' outcomes are: a smaller standard deviation on a given measure would indicate that participants had more similar outcomes. Standard deviations for highest grade completed are not included in Dynarski et al. (1998) and were reported to the WWC by the study authors.
3. Positive differences and effect sizes favor the intervention group; negative differences and effect sizes favor the comparison group. Means from Dynarski et al. (1998) are estimated using regression models that control for baseline characteristics.
4. For an explanation of the effect size calculation, see [Technical Details of WWC-Conducted Computations](#).
5. Statistical significance is the probability that the difference between groups is a result of chance rather than a real difference between the groups. The WWC obtained additional information on the level of statistical significance of this difference from the study authors.
6. The improvement index represents the difference between the percentile rank of the average student in the intervention condition and that of the average student in the comparison condition. The improvement index can take on values between –50 and +50, with positive numbers denoting results favorable to the intervention group.
7. The level of statistical significance was reported by the study authors or, where necessary, calculated by the WWC to correct for clustering within classrooms or schools (corrections for multiple comparisons were not done for findings not included in the overall intervention rating). For an explanation about the clustering correction, see the [WWC Tutorial on Mismatch](#). See [Technical Details of WWC-Conducted Computations](#) for the formulas the WWC used to calculate statistical significance. In the case of Dynarski et al. (1998), no corrections for clustering or multiple comparisons were needed.

Appendix A5.1 *Twelve Together* rating for the staying in school domain

The WWC rates an intervention's effects for a given outcome domain as positive, potentially positive, mixed, no discernible effects, potentially negative, or negative.¹

For the outcome domain of staying in school, the WWC rated *Twelve Together* as having potentially positive effects. It did not meet the criteria for positive effects because only one study of *Twelve Together* met evidence standards with reservations. The remaining ratings (mixed effects, no discernible effects, potentially negative effects, and negative effects) were not considered because *Twelve Together* was assigned the highest applicable rating.

Rating received

Potentially positive effects: Evidence of a positive effect with no overriding contrary evidence.

- Criterion 1: At least one study showing a statistically significant or substantively important *positive* effect.

Met. The one study of *Twelve Together* that passed WWC evidence screens found a substantively important positive effect on staying in school (effect size > 0.25).

- Criterion 2: No studies showing a statistically significant or substantively important *negative* effect and fewer or the same number of studies showing *indeterminate* effects than showing statistically significant or substantively important *positive* effects.

Met. No studies found statistically significant or substantively important negative effects in this domain.

Other ratings considered

Positive effects: Strong evidence of a positive effect with no overriding contrary evidence.

- Criterion 1: Two or more studies showing statistically significant *positive* effects, at least one of which met WWC evidence standards for a strong design.

Not met. *Twelve Together* had only one study meeting WWC evidence standards.

- Criterion 2: No studies showing statistically significant or substantively important *negative* effects.

Met. No studies found statistically significant or substantively important negative effects in this domain.

1. For rating purposes, the WWC considers the statistical significance of individual outcomes and the domain level effects. The WWC also considers the size of the domain level effects for ratings of potentially positive or potentially negative effects. See the [WWC Intervention Rating Scheme](#) for a complete description.

Appendix A5.2 *Twelve Together* rating for the progressing in school domain

The WWC rates an intervention's effects for a given outcome domain as positive, potentially positive, mixed, no discernible effects, potentially negative, or negative.¹

For the outcome domain of progressing in school, the WWC rated *Twelve Together* as having no discernible effects. It did not meet the criteria for other ratings (positive effects, potentially positive effects, mixed effects, potentially negative effects, and negative effects) because the one study that met WWC evidence standards did not show statistically significant or substantively important effects.

Rating received

No discernible effects: No affirmative evidence of effects.

- Criterion 1: None of the studies shows a statistically significant or substantively important effect, either *positive* or *negative*.

Met. No studies found statistically significant or substantively important effects in this domain.

Other ratings considered

Positive effects: Strong evidence of a positive effect with no overriding contrary evidence.

- Criterion 1: Two or more studies showing statistically significant *positive* effects, at least one of which met WWC evidence standards for a strong design.

Not met. *Twelve Together* had only one study meeting WWC evidence standards.

- Criterion 2: No studies showing statistically significant or substantively important *negative* effects.

Met. No studies found statistically significant or substantively important negative effects in this domain.

Potentially positive effects: Evidence of a positive effect with no overriding contrary evidence.

- Criterion 1: At least one study showing a statistically significant or substantively important *positive* effect.

Not met. No studies found statistically significant or substantively important positive effects in this domain.

- Criterion 2: No studies showing a statistically significant or substantively important *negative* effect and fewer or the same number of studies showing *indeterminate* effects than showing statistically significant or substantively important *positive* effects.

Met. No studies found statistically significant or substantively important negative effects in this domain.

Mixed effects: Evidence of inconsistent effects as demonstrated through either of the following criteria.

- Criterion 1: At least one study showing a statistically significant or substantively important *positive* effect, and at least one study showing a statistically significant or substantively important *negative* effect, but no more such studies than the number showing a statistically significant or substantively important *positive* effect.

Not met. No studies found statistically significant or substantively important effects, either positive or negative, in this domain.

- Criterion 2: At least one study showing a statistically significant or substantively important effect, and more studies showing an *indeterminate* effect than showing a statistically significant or substantively important effect.

Not met. No studies found statistically significant or substantively important effects in this domain.

(continued)

Appendix A5.2 *Twelve Together* rating for the progressing in school domain (continued)

Potentially negative effects: Evidence of a negative effect with no overriding contrary evidence

- Criterion 1: At least one study showing a statistically significant or substantively important *negative* effect.

Not met. No studies found statistically significant or substantively important negative effects in this domain.

- Criterion 2: No studies showing a statistically significant or substantively important *positive* effect, or more studies showing statistically significant or substantively important *negative* effects than showing statistically significant or substantively important *positive* effects.

Met. No studies found statistically significant or substantively important positive effects in this domain.

Negative effects: Strong evidence of a negative effect with no overriding contrary evidence.

- Criterion 1: Two or more studies showing statistically significant *negative* effects, at least one of which met WWC evidence standards for a strong design.

Not met. No studies found statistically significant negative effects in this domain.

- Criterion 2: No studies showing statistically significant or substantively important *positive* effects.

Met. No studies found statistically significant or substantively important positive effects in this domain.

1. For rating purposes, the WWC considers the statistical significance of individual outcomes and the domain level effects. The WWC also considers the size of the domain level effects for ratings of potentially positive or potentially negative effects. See the [WWC Intervention Rating Scheme](#) for a complete description.